Measures of Disease Frequency II

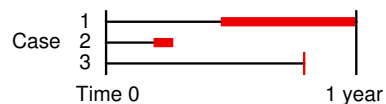## Seattle Epidemiology and Biostatistics Summer Session
June, 2004

Prevalence and Length-Biased Sampling

Prevalence and duration of disease

- Duration of disease can vary among individuals
- Example: coronary heart disease

Case 1 2 3
Time 0   1 year

- Probability of being in diseased state when a prevalence survey is done is proportional to *duration of disease*
- Prevalent cases are thus skewed toward more chronic forms of disease

Implications for case-control studies

- Case-control studies compare frequency of an exposure between disease cases and non-diseased persons
- If prevalent cases used, associations may be due to effect of exposure on either:
  - Risk of *developing* disease, *or*
  - *Chronicity* of disease
- To focus on risk factors for *developing* disease, using incident cases preferred

Implications for screening

- Screening program is like a prevalence survey
- Cases with a long asymptomatic phase more likely to be detected
- These may be slowly progressive cases with generally better prognosis
- Survival may thus appear better in screen-detected cases than in other cases, even if early detection does not prolong life

Rates in Total Population and Subpopulations

Subpopulations

Populations can be divided in many ways into *mutually exclusive and collectively exhaustive* subgroups—e.g.:
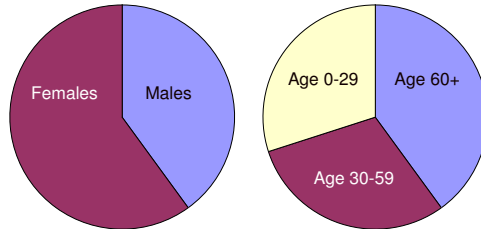
Subgroup-specific rates

Disease frequency can be measured in a subgroup just as for full population—e.g.:

Subgroup	Cases	No. at risk	Cumulative incidence
Males	40	400	10%
Females	90	600	15%
TOTAL	130	1000	13%

Rate in full population as a weighted average

Using data from slide 8:

$$
\begin{aligned}
CI_{total} &= \frac{130}{1000} \\
&= \frac{40+90}{1000} \\
&= \frac{40}{1000} + \frac{90}{1000} \\
&= \left(\frac{400}{1000}\right) \cdot \left(\frac{40}{400}\right) + \left(\frac{600}{1000}\right) \cdot \left(\frac{90}{600}\right) \\
&= w_{males} \cdot CI_{males} + w_{females} \cdot CI_{females}
\end{aligned}
$$

The *Weighted Average Rule*

- Rate in full population is a *weighted average* of the rates in its component subgroups

- Weight for each subgroup-specific rate is the *proportion of the full population that belongs to that subgroup*

  $R_{total} = \sum_i w_i \cdot R_i$

 where w = proportion of full population in subgroup i

- Here, R can be *any* measure of disease frequency that takes the form of a fraction

Incidence Rate and Cumulative Incidence

Scenario

- Closed population with all members initially at risk

- Incidence rate (IR) of a non-recurrent disease remains constant over time

- As new cases occur, the population still at risk declines over time

- Hence the *number* of new cases per unit of time also declines over time

What happens over time
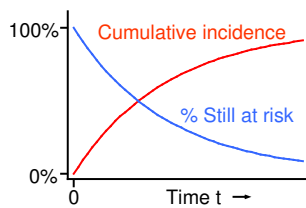
- Proportion remaining at risk declines *exponentially* over time
- Cumulative incidence increases over time, but with steadily decreasing slope
- $CI = 1 - e^{-IR \cdot t}$

Example

- Say that incidence rate (IR) of HPV infection in sexually active college women is 25 cases per 100 woman-years
- What proportion of women will have been infected with HPV at end of each year?

Year	Formula	Cumulative incidence of infection
1	$1 - e^{(-0.25 \cdot 1)}$	.22 = 22%
2	$1 - e^{(-0.25 \cdot 2)}$	.39 = 39%
3	$1 - e^{(-0.25 \cdot 3)}$	.53 = 53%
4	$1 - e^{(-0.25 \cdot 4)}$	.63 = 63%

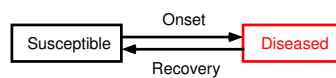*Prevalence, Incidence, and Duration of Disease*

The scenario

A recurrent disease, with only a *susceptible* and a *diseased* state

Susceptible → Onset → Diseased
Diseased → Recovery → Susceptible
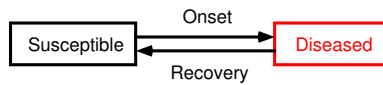
(View simulation)

What happens over time—1

Susceptible → Onset → Diseased
Diseased → Recovery → Susceptible
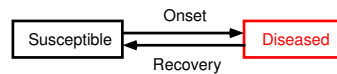
- No. of incident cases per unit time depends on:
  - No. of susceptibles
  - Incidence rate
- No. of recovering cases per unit time depends on
  - No. of prevalent disease cases
  - "Recovery" rate—like incidence rate, but in opposite direction

What happens over time—2

- System tends toward equilibrium over time
    - If the 2 flows are unequal, one compartment grows while the other shrinks. . .
    - . . . which, in turn, tends to equalize the 2 flows
- Can show (see text) that at equilibrium:

$$\text{Mean duration of disease } (\overline{d}) = 1/(\text{recovery rate})$$

$$\text{Prevalence} \approx (\text{Incidence rate}) \cdot \overline{d}$$

Example: pregnancy

Birth control use	Incidence rate* of pregnancy	Duration (years)	Predicted prevalence (approx.)
None	8	0.75	6%
50% use OCs	4	0.75	3%
50% have abortion at 3 months	8	0.50	4%

*Pregnancies per 100 woman-years

Caveats

- Assumes that incidence rate, disease duration have been stable long enough for system to achieve equilibrium
- In real life, this assumption is often not satisfied
    - *IR* can vary due to changes in population exposures over time
    - $\overline{d}$ can vary due to changes in treatment patterns
- Hence rule best regarded as a conceptual aid; don't expect exact agreement with real data

Mortality, Incidence,
and Case Fatality

Mortality $\approx$ incidence $\times$ case fatality

- $\begin{array}{c}\text{Risk of}\\\text{dying of}\\\text{a disease}\end{array} \approx \left(\begin{array}{c}\text{Risk of}\\\text{getting the}\\\text{disease}\end{array}\right) \times \left(\begin{array}{c}\text{Risk of dying}\\\text{from the disease}\\\text{if you get it}\end{array}\right)$

- In symbols:
    - $MR \approx IR \times CFR$
    - $CM \approx CI \times CFR$

- Strict equality requires stable incidence and case fatality long enough to reach a steady state

Example: uterine rupture in childbirth

- Over 10 years in Southwestern Nigeria:

Deliveries:	16,683
Uterine ruptures:	61
Deaths from uterine rupture:	13

- From these data:

Cumulative incidence:	61/16,683	= 3.66 per 1000
Case fatality:	13/61	= .213
Cumulative mortality:	13/16,683	= 0.78 per 1000

    (Source: *Singapore Med J* 2004; 45:113–6)

Introduction to Epidemiologic Methods — Summer, 2004
Discussion Questions: Measures of Disease Frequency—2

1. In 1980, 31.9% of all deaths among persons age 45–64 years in the U.S. were due to cancer. In 1999, this figure had risen to 34.6%. Assume that (1) the accuracy of cause-of-death recording remained constant, and (2) the increase was well beyond what chance fluctuations alone could explain. Would it be correct to state that the risk of dying of cancer rose slightly for U.S. adults in this age range from 1980 to 1999?

2. In a statewide survey, the prevalence of current smoking among military veterans age 30–59 years was found to be 32% overall—33% among male veterans and 24% among female veterans. About 95% of veterans in this age group were men.

    By comparison, among non-veterans of similar age, the prevalence of current smoking was only 26%—about 6 percentage points lower than in veterans. However, the non-veteran population was about a 50:50 mix of men and women.

    Suppose you are interested in determining whether the prevalence of smoking really differs according to veteran status, apart from their gender differences. Can you project what the prevalence of current smoking would be in a hypothetical population of veterans that consisted of a 50:50 mix of male and female veterans?

3. A study of mortality among homeless men in New York City identified 332 middle-aged men in homeless shelters on a particular night in 1987. Through linkage to the National Death Index, it was determined that 43 (13.1%) of these men died during the following 7 years.

    Vital statistics for New York City showed that the mortality rate for comparably aged males at that time was about 9.1 deaths per 1000 person-years. If this mortality rate had applied to the 332 homeless men in the sample, how many deaths would have been expected during the 7-year period?